Click Here for Full Article

Joint Bayesian model selection and parameter estimation of the generalized extreme value model with covariates using birth-death Markov chain Monte Carlo

Salaheddine El Adlouni[1] and Taha B. M. J. Ouarda[2]

Received 10 August 2007; revised 11 November 2008; accepted 11 March 2009; published 5 June 2009.

[1] This paper describes Bayesian estimation of the parameters of the generalized extreme value (GEV) model with covariates. For this model the parameters of the GEV distribution are functions of covariates, allowing for dependent parameters and/or trends. A Markov chain Monte Carlo (MCMC) algorithm is generally used to estimate the posterior distributions of the parameters in a Bayesian framework. In this paper, the birth-death MCMC (BDMCMC) procedure is developed in order to carry out both parameter estimation and Bayesian model selection. The BDMCMC methods allow the jump between models of different dimensions. The general algorithm consists of two types of sampling steps. The first one involves dimension-changing moves, and the second is conditional on a fixed model. Parameters are estimated in a fully Bayesian framework, and the model is selected by the length of time that the MCMC chain remains in that model. Real and simulated data sets illustrate the usefulness of the proposed methodology.

Citation: El Adlouni, S., and T. B. M. J. Ouarda (2009), Joint Bayesian model selection and parameter estimation of the generalized extreme value model with covariates using birth-death Markov chain Monte Carlo, *Water Resour. Res.*, *45*, W06403, doi:10.1029/2007WR006427.

1. Introduction and Review

[2] Extreme value analysis allows the interpretation of past records and the inference about future probabilities of occurrence of extreme events, such as floods, extreme rainfalls, or high wind gusts. Extreme values are often represented by the maximum value of the variable of interest over a given time period, such as a year. Extreme value theory indicates that these maxima can generally be described by one of the three extreme value distributions that can be represented by the generalized extreme value (GEV) distribution [e.g., *Jenkinson*, 1955]. To estimate the parameters of the GEV distribution, several approaches were proposed in the literature to avoid the computational problems related to the maximum likelihood approach especially for small samples. *Coles and Dixon* [1999] proposed the penalized maximum likelihood method, which retains the modeling flexibility of the maximum likelihood estimator and improves its small-sample properties. *Martins and Stedinger* [2000] developed the generalized maximum likelihood (GML) method for hydrometeorological series. This approach was also investigated by *Park* [2005], who studied the optimal selection of the hyperparameters. Recently, *Zhang* [2007] proposed the likelihood moment estimator, which is computationally simple and possesses asymptotic efficiency.

[3] There are three fundamental assumptions for classical frequency analysis to provide useful engineering design values. The observations should be independent and the data series should be stationary and homogeneous to ensure that the statistical inference will be valid during the projected life span of the engineering structure. There is, however, mounting evidence that such assumptions are not always valid. Indeed, statistically significant trends have been identified in observed historical extreme events of different hydroclimatological series [*Intergovernmental Panel on Climate Change* (IPCC), 2007] in different parts of the world, and climate extremes will likely change in the future [e.g., *Kharin and Zwiers*, 2005]. The reality of nonstationary hydrometeorological extremes needs to be properly addressed because the GEV model with constant parameters may no longer be valid under nonstationary conditions [*Leadbetter et al.*, 1983]. In this case, a new definition should be given to return period events for risk assessment. Indeed, the common notion of "return period" is no longer appropriate in a nonstationary framework. Recently, *Stefanakos and Athanassoulis* [2006] developed a new definition of the return period notion, based on the mean number of upcrossings (MENU) of a given level. The MENU approach can be used to give an operational computation of the return period event for nonstationary hydrometeorological variables. When the covariate is not time, risk assessment can be carried out by considering the worst-case scenario, which may occur toward the end of the lifetime of the structure.

[4] The use of the models with covariates makes it possible to combine the effect of other variables to classical frequency analysis models [*Zhang et al.*, 2004; *Clarke*, 2002]. Two principal problems are related to the implementation of such models. The first one involves the complexity

[1]Institut National de Statistique et d'Économie Appliquée, Rabat, Morocco.
[2]Hydro-Quebec/Natural Sciences and Engineering Research Council of Canada, University of Quebec, Quebec, Ontario, Canada.

0043-1397/09/2007WR006427$09.00

of parameter estimation techniques, and the second one concerns model selection. Indeed, several competing models can be considered, and hence robust techniques for model selection need to be developed.

[5] *Coles* [2001] provided a general description of the covariate approach for the generalized extreme value distribution (GEV) and presented the maximum likelihood (ML) estimation method. *El Adlouni et al.* [2007] extended the generalized maximum likelihood (GML) method, developed by *Martins and Stedinger* [2000] for the stationary case, to the case with covariates. In the GML method, the shape parameter has a beta distribution as prior defined on the interval $[-0.5, 0.5]$. The use of such prior, for hydrometeorological series, makes it possible to avoid the instability and convergence problems recorded for the ML method. The GML method can be generalized and the prior for the shape parameter can be used in a fully Bayesian framework. Indeed, the Bayesian approach gives a general and elegant way to integrate any additional information.

[6] Until the mid-1990s, the use of the Bayesian approach, in practice, was restricted to some simple problems with explicit solutions. However, computational difficulties related to the estimation of the Bayesian posterior distribution were overcome by using Markov chain Monte Carlo methods (MCMC). The use of the Bayesian framework allows the integration of any additional information in the inference process. *Reis and Stedinger* [2005] used a Bayesian framework to integrate regional and historical information. *Coles and Tawn* [1996] and *Stephenson and Tawn* [2004] used elicitation techniques based on quantile differences for the GEV distribution.

[7] In the Bayesian framework, common methods for model selection are based on posterior predictive distributions or Bayes factors. Because the Bayes factors are often difficult to compute, a good alternative is to adopt an approximation to the Bayesian information criterion (BIC) or the deviance information criterion (DIC). The DIC was first formulated for generalized linear models. *Celeux et al.* [2006] discussed alternative representations of the DIC for latent variables. Methods based on the posterior predictive p-values and conditional p-values are also emerging as popular measures of model fit [*Bayarri and Berger*, 2000; *Pérez and Berger*, 2002; *Aitkin et al.*, 2005]. Other techniques are based on comparative parameter estimation including distance measures such as entropy distance or Kullback-Leibler divergence.

[8] Another class of methods is based on enlarging the parameter space to include a maximum number of models of interest [*George and McCulloch*, 1993]. A very popular alternative is the reversible jump Markov chain Monte Carlo (RJMCMC), also called transdimensional MCMC [*Green*, 1995]. In the RJMCMC technique, the MCMC algorithm is enlarged to allow jumps between models with different parameter space dimensions. These methods are largely used for mixtures of distributions with an unknown number of components [*Richardson and Green*, 1997; *Robert et al.*, 2000; *Dellaportas and Papageorgiou*, 2006]. *Richardson and Green* [1997] suggest an additional Metropolis-Hastings step that involves proposals for the "birth" of a new component or "death" of an existing component. These moves require jumps between parameters of models of different dimensions. For example, for the GEV model with covariates, the MCMC algorithm should offer the possibility to jump from a linear dependence on covariates to a quadratic case or from a model with covariates in the location parameter to models with both location and scale parameters as functions of covariates. Recently, *Ribatet et al.* [2007] used the RJMCMC approach for regional flood frequency analysis. The proposed approach allows jumps from a general regional generalized Pareto (GPD) to a GPD model with a fixed-shape parameter derived from a homogenous region.

[9] In the present study, we develop a RJMCMC algorithm for parameter estimation and model selection for the GEV models with covariates. Parameters are estimated in a fully Bayesian framework, and the model is selected by the length of time that the MCMC chain remains in that model.

[10] The remainder of this paper is organized as follows. In section 2, we introduce the GEV model with covariates. Section 3 deals with Bayesian parameter estimation and model selection. Section 4 describes the RJMCMC algorithms corresponding to moves with fixed model and jumps between models with different space dimensions. In section 5 we present two illustrative examples: The first one is on simulated data to illustrate the method's potential, and the second one is on observed annual maximum precipitation data at the Tehachapi station in California (United States). Conclusions are given in section 6.

2. The GEV Model With Covariates

[11] The three types of extreme value distributions can be combined to form a single parametric family that is the generalized extreme value (GEV) distribution [*Jenkinson*, 1955]. The cumulative distribution function of the GEV distribution is

$$\begin{aligned} F_{GEV}(x) &= \exp\left[-\left(1-\frac{\kappa}{\alpha}(x-\mu)\right)^{1/\kappa}\right] \quad \kappa \neq 0 \\ &= \exp\left[-\exp\left(-\frac{(x-\mu)}{\alpha}\right)\right] \quad \kappa = 0 \end{aligned} \tag{1}$$

where $\mu + \alpha/\kappa \leq x < +\infty$ when $\kappa < 0$ (corresponding to the Fréchet distribution), $-\infty < x < +\infty$ when $\kappa = 0$ (Gumbel), and $-\infty < x \leq \mu + \alpha/\kappa$ when $\kappa > 0$ (Weibull). The $\mu(\in\mathbb{R})$, $\alpha(>0)$, and $\kappa(\in\mathbb{R})$ are the location, the scale, and the shape parameters, respectively.

[12] In the case of the model with covariates, the parameters depend on other variables such as time: $GEV(\mu_t, \alpha_t, \kappa_t)$ [*Coles*, 2001]. To ensure a positive value for the scale parameter, a transformation such that $\varphi_t = \log(\alpha_t)$ is used when estimating the parameters. We assume that the location parameter μ_t is a function of n_μ covariates $U = (U_1\ U_2 \ldots U_{n_\mu})'$. Let $\beta = (\beta_1\ \beta_2 \ldots \beta_{n_\mu})'$ be the vector of corresponding parameters. In the case of linear dependence we have

$$\mu_t = U'(t)\beta = \sum_{i=1}^{n_\mu} \beta_i U_i(t). \tag{2}$$

For the scale parameter α_t, let $V = (V_1\ V_2 \ldots V_{n_\alpha})'$ be the vector of covariates. We have

$$\varphi_t = \log(\alpha_t) = V'(t).\delta = \sum_{i=1}^{n_\alpha} \delta_i V_i(t), \tag{3}$$

where $\delta = (\delta_1\ \delta_2 \ldots \delta_{n_\alpha})'$ are the corresponding parameters. The same applies to the shape parameter κ_t,

$$\kappa_t = W'(t).\gamma = \sum_{i=1}^{n_\kappa} \gamma_i W_i(t), \quad (4)$$

where $W = (W_1\ W_2 \ldots W_{n_\kappa})'$ are the covariates and $\gamma = (\gamma_1\ \gamma_2 \ldots \gamma_{n_\kappa})'$ are the corresponding parameters.

[13] For the GEV model with covariates, the likelihood function for a given sample $\underline{x}'_n = \{x_1, \ldots, x_n\}$ is

$$L_n = \prod_{t=1}^{n} f(x_t | \mu_t, \varphi_t, \kappa_t), \quad (5)$$

where f is the probability density function (pdf) of the GEV distribution.

[14] The shape parameter is taken to be constant ($\kappa_t = \kappa$). For the location and scale parameters, dependence is linear and the number of covariates is restricted to $1 \le n_\mu \le n_\mu(\max)$ and $1 \le n_\alpha \le n_\alpha(\max)$ in order to limit the number of parameters to be estimated. However, several models can still be considered given particular values of n_μ and n_α. Here are some of these models:

[15] 1. $GEV_{1,1}(\mu, \alpha, \kappa)$ is the classic model with all parameters being constant: $\mu_t = \mu$, $\alpha_t = \alpha$ et $\kappa_t = \kappa$. In this case, $n_\mu = n_\alpha = 1$.

[16] 2. $GEV_{2,1}(\mu_t = \beta_1 + \beta_2 Y_t, \alpha, \kappa)$ is the homoscedastic model, and the location parameter is a linear function of one covariate Y_t ($n_\mu = 2$, $U(t) = (U_1(t) = 1\ U_2(t) = Y_t)$, $n_\alpha = 1$, and $V_1 = 1$).

[17] 3. In the $GEV_{2,2}(\mu_t = \beta_1 + \beta_2 Y_t, \alpha_t = \exp(\delta_1 + \delta_2 Y_t), \kappa)$ model, the location and scale parameters are function of the covariate Y_t. This model is recommended when the covariate is time $Y_t = t$, because trends are usually observed at the same time in the location and scale parameters ($n_\mu = 2$, $U(t) = (U_1(t) = 1\ U_2(t) = Y_t)$, $n_\alpha = 2$, $V(t) = (V_1(t) = 1\ V_2(t) = Y_t)$).

[18] 4. In the $GEV_{3,2}(\mu_t = \beta_1 + \beta_2 Y_t + \beta_3 Y_t^2, \alpha_t = \exp(\delta_1 + \delta_2 Y_t), \kappa)$ model, the location is a quadratic function of the covariate Y_t and the scale parameter is a linear function of the same covariate ($n_\mu = 3$, $U(t) = (U_1(t) = 1\ U_2(t) = Y_t\ U_3(t) = Y_t^2)$, $n_\alpha = 2$, $V(t) = (V_1(t) = 1\ V_2(t) = Y_t)$).

[19] In the same manner, model $GEV_{1,2}$ can be defined as the model with a constant location parameter and a scale parameter that is a linear function of the covariate. $GEV_{3,1}$ is the model with a location expressed as a quadratic function of the covariate and a constant scale parameter. Other models can be defined using a vector of covariates. These models will be presented in general form in the rest of the paper.

3. Parameter Estimation and Model Selection

[20] To estimate the parameters of the GEV model with covariates, the maximum likelihood method is the most commonly used approach [*Coles*, 2001]. More recently, the generalized maximum likelihood (GML) method, developed originally for the classical model (GEV), was extended to the case with covariates [*El Adlouni et al.*, 2007]. The GML method improves the ML method considerably and avoids some computing problems related to the ML maximization. For the model choice problem, several models are fitted to the data and then several tests and criteria can be used for model comparison (such as the deviance statistics). In this paper, we propose a joint Bayesian parameter estimation and model selection approach using the reversible jump Markov chain Monte Carlo (RJMCMC) technique.

[21] Let $\Sigma = \{GEV_{n_\mu, n_\alpha};\ n_\mu = 1, \ldots, n_\mu(\max),\ n_\alpha = 1, \ldots, n_\alpha(\max)\}$ be the set of all considered models to fit the observed data set $\underline{x}'_n = (x_1, \ldots, x_n)$. For a given model, GEV_{n_μ}, n_α, the parameter space dimension is $(n_\mu + n_\alpha + 1)$. Bayesian inference is based on the parameter posterior distribution which is proportional to the product of the prior distribution and the likelihood function. In some cases, the posterior distribution cannot be given in an explicit form, and the parameter estimators are obtained by simulation. In this section, we outline the general approach to implement the MCMC algorithm.

[22] For known n_μ and n_α, samples from the joint posterior distribution of the parameters are generated using the Metropolis-Hastings algorithm. For unknown n_μ and n_α, the parameter spaces have different dimensions and thus the reversible jump samplers [*Green*, 1995] are required. *Richardson and Green* [1997] proposed the reversible jump MCMC (RJMCMC) algorithm and used a birth-death move to change the number of components of the mixtures. On the basis of a continuous time birth-death process, *Stephens* [2000] proposes a birth-death MCMC (BDMCMC) algorithm. We resort to the BDMCMC algorithm to jump between model spaces with different dimensions because it is straightforward to implement in our context.

[23] The general algorithm is based on two types of sampling steps. The first one involves dimension-changing moves, while the other is conditional on a fixed model. The algorithm generates samples from the posterior distribution of the parameters for all considered models. This algorithm will be presented in section 4.

3.1. Parameter Estimation

[24] In reality, for the GEV model with covariates, "the number of things you don't know is one of the things you don't know" [*Green*, 2003]. Thus for unknown n_μ and n_α, the parameter vector to be estimated is $\Re = (n_\mu, n_\alpha, \beta^{(n_\mu)}, \delta^{(n_\alpha)}, \kappa)$. Note that the dimension of the parameter space depends on the selected model and is given by $d = n_\mu + n_\alpha + 3$.

3.1.1. Parameter Prior Distributions

3.1.1.1. Model Identifier n_μ and n_α

[25] In the Bayesian framework, parameter prior distributions represent all the anterior knowledge on the model independent of the collected data. Several sources of prior information are used in practice. In hydrology, historical and/or regional information constitute a good source for prior knowledge [*Reis and Stedinger*, 2005]. In this study we use a discrete uniform distribution ($DU(1,n(\max))$) as prior for the parameters n_μ and n_α. The prior probability density function for the parameter n_μ is given by

$$\pi(n_\mu = i) = \begin{cases} \frac{1}{n_\mu(\max)} & if \quad i = 1, \ldots, n_\mu(\max) \\ 0 & otherwise \end{cases}. \quad (6)$$

3.1.1.2. GEV With Covariates Model Parameters

[26] To specify the prior distributions of the rest of the parameters, several techniques are proposed in the literature. For the stationary GEV model, *Coles and Tawn* [1996]

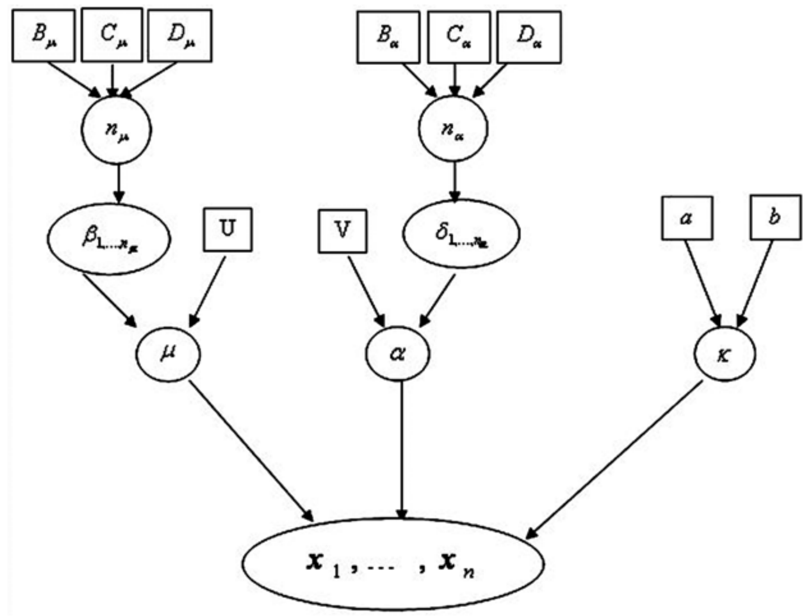

Figure 1. Directed acyclic graph (DAG) for the nonstationary GEV_{n_μ,n_α} model.

proposed a prior construction method based on the differences of given quantiles. Three quantiles should first be elicited and three independent priors for the differences are deduced. In this study, physical prior information for hydrometeorological data is used for the shape parameter of the GEV distribution. This prior is proposed by *Martins and Stedinger* [2000] to develop the GML method which was generalized for the GEV model with covariates by *El Adlouni et al.* [2007] In the GML approach, the prior distribution of the shape parameter is a beta distribution ($\kappa \sim \pi(\kappa) = Beta(p = 6, q = 9)$) defined on the interval [$a = -0.5$, $b = 0.5$]. The probability density function of this distribution is given by

$$f(x) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \frac{(x-a)^{p-1}(b-x)^{q-1}}{(b-a)^{p+q-1}}, \; a < x < b,$$

and the mean and the variance of this prior distribution are $E_\pi[\kappa] = -0.1$ and $Var_\pi[\kappa] = 0.12$.

[27] An important but reasonable assumption is that the parameters are independent, allowing us to define their priors separately. For the shape parameter, the same prior as the one used for the GML method is considered in a fully Bayesian framework. For all other parameters ($\beta^{(n_\mu)}$ and $\delta^{(n_\alpha)}$) the noninformative uniform prior is used. However, if we had additional knowledge from other sources regarding the scale and position parameters, we could include it, directly or by elicitation [*O'Hagan*, 2006].

[28] The joint prior distribution considered in this study for the vector of the parameters $\Re = (n_\mu, n_\alpha, \beta^{(n_\mu)}, \delta^{(n_\alpha)}, \kappa)$ is given by

$$\pi(\Re) \propto \pi(n_\mu)\ \pi(n_\alpha)\ \pi(\kappa). \tag{7}$$

3.1.2. Posterior Distribution Assessment

[29] For a given GEV_{n_μ,n_α} model and using the Bayes theorem, the posterior distribution is proportional to the product of the likelihood function $L_n(\underline{x}_n|\Re)$ (equation (5)) and the prior distribution (equation (7)),

$$\pi(\Re|\underline{x}_n) \quad \propto \quad L_n(\underline{x}_n|\Re)\ \pi(\Re). \tag{8}$$

Bayesian inference is based on this posterior distribution. However, marginal posterior distributions cannot be deduced explicitly from equation (8), and computational techniques should be used. To solve such Bayesian problems, we use a Metropolis-Hastings algorithm. A major advantage of all MCMC techniques is that the posterior distributions of the parameters and quantiles (or any function of the parameters) are easily evaluated with their empirical distributions. Quantities of interest, for example, in hydrological studies, concern point estimates of the quantiles, their credibility intervals, and predictive distributions. A directed acyclic graph (DAG), which defines the structure of the Bayes model, is presented in Figure 1. The rest of the parameters used in the DAG are presented in section 3.2.

[30] For model selection, we focus on the (marginal) maximum a posteriori (MAP) estimators of n_μ and n_α, since they are discrete random variables. The MAP estimator, which minimizes the risk function r corresponding to the $0-1$ cost function, is defined by

$$\hat{n}_\theta = \arg\max_r \pi(r|\underline{x}_n) \tag{9}$$

with $\theta = \mu, \alpha$.

[31] For the rest of the parameters, which have continuous supports, we consider the standard Bayesian estimator which corresponds to the minimum mean square estimator (MMSE). The MMSE estimator of the parameter θ minimizes the quadratic cost function and is defined by

$$\hat{\theta} = E[r|\underline{x}_n], \tag{10}$$

where $\theta = \beta^{(n_\mu)}, \delta^{(n_\alpha)}, \kappa$.

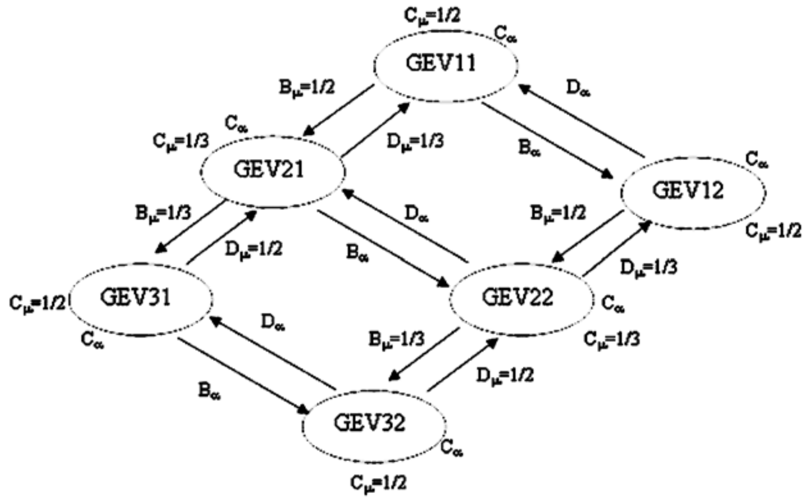

Figure 2. Illustration of model transitions for $n_\mu(\max) = 3$ and $n_\alpha(\max) = 2$.

3.2. Model Selection

[32] As mentioned above, the proposed approach uses RJMCMC to transit between models with different parameter space dimensions. Allowed transitions and corresponding probabilities concern closed states and are given by

$$\begin{cases} B_\mu = \text{probability of jump from} & n_\mu \longrightarrow n_\mu + 1 \\ B_\alpha = \text{probability of jump from} & n_\alpha \longrightarrow n_\alpha + 1 \\ D_\mu = \text{probability of jump from} & n_\mu \longrightarrow n_\mu - 1 \\ D_\alpha = \text{probability of jump from} & n_\alpha \longrightarrow n_\alpha - 1 \\ C_\mu = \text{probability of jump from} & n_\mu \longrightarrow n_\mu \\ C_\alpha = \text{probability of jump from} & n_\alpha \longrightarrow n_\alpha \end{cases}. \tag{11}$$

[33] For all components, the jump to a closed model is uniform distributed. Figure 2 illustrates transition possibilities for $n_\mu(\max) = 3$ and $n_\alpha(\max) = 2$. In this special case the prior transitions are presented in Figure 2. For all these cases, $B_\alpha = D_\alpha = C_\alpha = 1/2$, given that for the scale parameter the birth and death are not possible simultaneously.

[34] Note that jumps related to the location parameter with different space dimensions are independent from those of the scale parameter. In what follows, we present all possible jumps and corresponding accept/reject posterior probabilities.

4. Reversible Jump MCMC Algorithm

4.1. General MCMC Method

[35] The Markov chain Monte Carlo (MCMC) method is a powerful tool for Bayesian estimation. MCMC sampling was first introduced by *Metropolis et al.* [1953] to integrate over high dimensional probability distributions to make inference about model parameters. In Bayesian inference we are interested in finding the joint posterior distribution of the parameters. The difficulty is that the posterior distribution is typically found by multidimensional integration, which is only feasible for small-scale problems and hence many problems become intractable. When the full-conditional densities are of standard form, including their normalizing constants, posterior sampling is usually done via the Gibbs sampler. If however, one or more full-conditional densities are not of standard form and have intractable normalizing constants, as it is the case for the studied model, posterior sampling is usually conducted via the Metropolis-Hastings (MH) algorithm. The Metropolis-Hastings sampling involves, for each component θ_i of the parameter vector $\Re$ drawn from a "proposal" distribution, deciding whether to keep the current value or instead keep this proposed value. Principal steps of the MH algorithm, to draw a sample from a given distribution π, can be summarized as follows: (1) Choose a proposal distribution Q. (2) Given the current state ξ, generate ξ^* from $Q(.|\xi)$. (3) Accept ξ^* with probability

$$\rho(\xi, \xi^*) = \min\left\{1, \frac{\pi(\xi^*)}{\pi(\xi)} \frac{Q(\xi|\xi^*)}{Q(\xi^*|\xi)}\right\}.$$

Under some regularity conditions (the chain is irreducible and aperiodic) the distribution of interest π is the stationary distribution of the constructed chain (which corresponds to the posterior distribution given in equation (8) in this case). A detailed description of these algorithms is given by *Robert and Casella* [2004].

4.2. RJMCMC Algorithm

[36] The MCMC approach, as given in general form, presents some difficulties when the parameter space E is a union of subspaces with different dimensions. Suppose for simplicity that $E = \mathbb{R}^{k_1} \cup \mathbb{R}^{k_2}$ with $k_1 < k_2$. To define the proposal distribution Q, we should define the jumping rule from $\mathbb{R}^{k_1}$ to $\mathbb{R}^{k_2}$, then from $\mathbb{R}^{k_2}$ to $\mathbb{R}^{k_1}$, in order to build an irreducible and aperiodic chain. *Green* [1995] proposed an elegant solution to this problem. Given the current state ξ in $\mathbb{R}^{k_1}$, we generate $k_2 - k_1$ components to have a vector of $\mathbb{R}^{k_2}$. Then we define a bijective transformation in $\mathbb{R}^{k_2}$ which gives the state ξ^*. In the last step, a transformation function g, usually deterministic, should be defined to move from ξ^* to ξ. These transformations allow the definition of the

proposal distribution Q, and the accept/reject probability is given by

$$\rho(\xi,\xi^*) = \min\left\{1, \frac{\pi(\xi^*)}{\pi(\xi)}\frac{Q(\xi|\xi^*)}{Q(\xi^*|\xi)}|J(\xi,\xi^*)|\right\}. \qquad (12)$$

This probability is the same as presented for the general MCMC algorithms with an additional factor J, which corresponds to the Jacobian of the transformation function.

[37] In our case, the parameter space is given by

$$E = \bigcup_{\substack{n_\mu=1,\ldots n_\mu(\max)\\ n_\alpha=1,\ldots n_\alpha(\max)}} \mathbb{R}^{n_\mu}\times\mathbb{R}^{n_\alpha}\times[-0.5;0.5].$$

As presented in section 4.2, a jump in the location parameter space corresponds to a birth or a death. The same transformations are available for the scale parameter.

4.3. Birth-Death MCMC Algorithm

[38] Here we present a birth-death MCMC (BDMCMC) algorithm to obtain a sample from the joint posterior distribution of the vector of parameters $\Re = (n_\mu, n_\alpha, \beta^{(n_\mu)}, \delta^{(n_\alpha)}, \kappa)$. The BDMCMC approach was introduced by *Stephens* [2000] for normal mixtures and is based on a birth-death process where the mixture size changes. As mentioned in section 3.2, transitions and corresponding probabilities concern closed states. Thus the BDMCMC is straightforwardly the most suitable in the present context. Algorithms presented here concern the location parameter; the same algorithms can be formulated for the scale parameter.

[39] Let n_μ be the current order of the GEV model with covariates, $\beta^{(n_\mu)} = (\beta_1,\ldots,\beta_{n_\mu})$ the vector of the parameters corresponding to the location parameter, and suppose that a jump from n_μ to $n_\mu + 1$ (birth) is selected using equation (11) and algorithm 1 (given hereinafter). A random $(n_\mu + 1)$ component is generated using proposal distribution g. Starting values of additional components are taken to be close to zero. We consider for g a triangular distribution on the interval $[a,b]$ with mode $c\in]a,b[$:

$$g(s) = \begin{cases} \dfrac{2(s-a)}{(b-a)(c-a)}, & \text{for } a\le s\le c \\ \dfrac{2(b-s)}{(b-a)(b-c)}, & \text{for } c< s\le b \end{cases}. \qquad (13)$$

In this study we use $a = -0.1$, $c = 0$, and $b = 0.1$.

[40] The generated component will be added to the vector $\beta^{(n_\mu)}$, and the proposed vector is

$$\beta^{(n_\mu+1)} = \left(\beta_1,\ldots,\beta_{n_\mu},s\right). \qquad (14)$$

The Jacobian of this transformation is equal to 1.

[41] In the case of a death transition, a jump from $(n_\mu + 1)$ to n_μ is selected. In this case the last component $\beta_{n_\mu} + 1$ is deleted and the first one is replaced by the mean on all covariate observations. This approach was suggested by an anonymous reviewer. The proposed vector is given by $\beta^{(n_\mu)} = \left(\beta_1 + \frac{1}{n}\sum_{i=1}^{n}\beta_{n_\mu+1}y_i,\ \beta_2,\ldots,\beta_{n_\mu}\right)$.

[42] For each transformation (birth or death) the acceptation/rejection probability can be given for both jumps:

$$p_B = \min(1, r_B) \ \text{ and } \ p_D = \min\left(1, r_B^{-1}\right) \qquad (15)$$

with

$$r_B = \frac{\pi\left(n_\mu+1, n_\alpha, \beta^{(n_\mu+1)}, \delta^{(n_\alpha)}|X_n\right)}{\pi\left(n_\mu, n_\alpha, \beta^{(n_\mu)}, \delta^{(n_\alpha)}|X_n\right)} \cdot \frac{Q\left(n_\mu, \beta^{(n_\mu)}|n_\mu+1,\ \beta^{(n_\mu+1)}\right)}{Q\left(n_\mu+1,\ \beta^{(n_\mu+1)}|n_\mu, \beta^{(n_\mu)}\right)} \qquad (16)$$

$$\begin{cases} Q\left(n_\mu, \beta^{(n_\mu)}|n_\mu+1,\ \beta^{(n_\mu+1)}\right) = D_\mu \\ Q\left(n_\mu+1,\ \beta^{(n_\mu+1)}|n_\mu, \beta^{(n_\mu)}\right) = B_\mu\, g\left(\beta^{(n_\mu+1)}\right) \end{cases}. \qquad (17)$$

The posterior distribution $\pi(.|X_n)$ is given by equation (8). Algorithms corresponding to birth and death transitions are given by:

[43] Algorithm 1 (Birth)

[44] 1. Generate $s \sim g$ (equation (13)).

[45] 2. Compute the probability of birth P_B (equation (15)).

[46] 3. If $u \sim U(0,1) \le P_B$, then $\beta^{(n_\mu+1)} = (\beta_1,\ldots,\beta_{n_\mu}, s)$, else the vector of parameters is $\beta^{(n_\mu)}$.

[47] Algorithm 2 (Death)

[48] 1. Delete the last component $\beta_{n_\mu+1}$ and set

$$\beta^{(n_\mu)} = \left(\beta_1 + \frac{1}{n}\sum_{i=1}^{n}\beta_{n_\mu+1}y_i, \beta_2,\ldots,\beta_{n_\mu}\right).$$

[49] 2. Compute the probability of death P_D (equation (15)).

[50] 3. If $u \sim U(0,1) \le P_D$, then

$$\beta^{(n_\mu)} = \left(\beta_1 + \frac{1}{n}\sum_{i=1}^{n}\beta_{n_\mu+1}y_i,\ldots,\beta_{n_\mu}\right),$$

else the vector of parameters is $\beta^{(n_\mu+1)}$.

[51] In the same manner, the probabilities of birth (P'_B) and death (P'_D) for the scale parameter are equivalent to equation (15) and are obtained by switching the roles of the parameters β and δ. Thus the birth and death algorithms corresponding to the scale parameter are given by algorithms 1′ and 2′:

[52] Algorithm 1′ (Birth)

[53] 4. Generate $s \sim g$ (equation (13)).

[54] 5. Compute the probability of birth P'_B.

[55] 6. If $u \sim U(0,1) \le P'_B$, then $\delta^{(n_\alpha+1)} = (\delta_1,\ldots,\delta_{n_\alpha}, s)$, else the vector of parameters is $\delta^{(n_\alpha)}$.

[56] Algorithm 2′ (Death)

[57] 4. Delete the last component $\delta_{n_\alpha+1}$ of $\delta^{(n_\alpha+1)}$ and set

$$\delta^{(n_\alpha)} = \left(\delta_1 + \frac{1}{n}\sum_{i=1}^{n}\delta_{n_\alpha+1}y_i, \delta_2,\ldots,\delta_{n_\alpha}\right).$$

[58] 5. Compute the probability of death P'_D.

[59] 6. If $u \sim U(0,1) \leq P'_D$, then

$$\delta^{(n_\alpha)} = \left(\delta_1 + \frac{1}{n}\sum_{i=1}^{n}\delta_{n_\alpha+1}y_i, \delta_2, \ldots, \delta_{n_\alpha}\right),$$

else the vector of parameters is $\delta^{(n_\alpha+1)}$.

4.4. Update of the MCMC Iterations

[60] Suppose that no jump is chosen. Consequently, there is no change in the dimension of the covariate parameters of the location parameter ($n_\mu \longrightarrow n_\mu$). The Metropolis-Hastings (MH) algorithm [*Metropolis et al.*, 1953; *Hastings*, 1970] is used to update the Markov chain iterations. Simulated realizations are generated from the posterior distribution using a single-component random-walk Metropolis algorithm with a Gaussian proposal density centered at the current states, ${\beta_{(i)}}^{(n_\mu)}$ and ${\delta_{(i)}}^{(n_\mu)}$, of the chain [*Gilks et al.*, 1996]. The variance of the proposal distribution is relatively small with respect to the parameter range allowing a displacement toward closest states. Let ${\beta_0}^{(n_\mu)}$ and ${\delta_0}^{(n_\mu)}$ be new candidates generated from the proposal distribution. The probability of acceptation/rejection is

$$p_U = \min(1, r_U), \tag{18}$$

where

$$r_U = \frac{\pi\left(n_\mu, n_\alpha, \beta_0^{(n_\mu)}, \delta_0^{(n_\alpha)}|X_n\right)}{\pi\left(n_\mu, n_\alpha, \beta_{(i)}^{(n_\mu)}, \delta_{(i)}^{(n_\alpha)}|X_n\right)}. \tag{19}$$

The algorithm corresponding to the update of the MCMC iterations, for the location parameters, is given as follows:

[61] Algorithm 3 (Update for β)

[62] 1. Compute the probability of update P_U (equation (18)).

[63] 2. If $u \sim U(0,1) \leq P_U$, then ${\beta_{(i+1)}}^{(n_\mu)} = {\beta_0}^{(n_\mu)}$, else the vector of the parameters is ${\beta_{(i+1)}}^{(n_\mu)} = {\beta_{(i)}}^{(n_\mu)}$.

[64] A similar algorithm corresponds to the vector of the parameters δ and is given as follows:

[65] Algorithm 3′ (Update for δ)

[66] 3. Compute the probability of Update P_U (equation (18)).

[67] 4. If $u \sim U(0,1) \leq P_U$, then ${\delta_{(i+1)}}^{(n_\alpha)} = {\delta_0}^{(n_\alpha)}$, else the vector of the parameters is ${\delta_{(i+1)}}^{(n_\alpha)} = {\delta_{(i)}}^{(n_\alpha)}$.

4.5. Practical Implementation of the BDMCMC Algorithm

[68] For the practical implementation of the BDMCMC algorithm several issues need special attention. The first one is the choice of starting values of all the parameters to be estimated. In theory, all generated Markov chains are irreducible by construction, meaning that all their states communicate (i.e., the probability to transfer from one state to another is non-null). Then the initial values should not influence the convergence of the chain to its stationary distribution. However, this choice influences the burn-in time and thus the length of the Markov chain to be generated. For all models considered in this study, the methodology adopted to define the starting value is as follows:

[69] 1. For the location parameters we choose the ordinary least squares estimators of the regression of the variable of interest when the covariates are taken as independent variables.

[70] 2. For the scale parameters, we consider the maximum likelihood estimator for the first scale parameter (δ_1) of the classical model and very low values for the other parameters.

[71] 3. For the shape parameter the maximum likelihood estimator is taken as starting value.

[72] The following is the algorithm in a schematic form. If the current state of the Markov chain is (n_μ, n_δ, GEV_{n_μ,n_δ}), then the proposed BDMCMC algorithm is as follows:

[73] BDMCMC Algorithm

[74] Step 1. Propose a transition from GEV_{n_μ,n_δ} to $GEV_{\underline{\underline{n}}\mu,n_\delta}$ with

[75] $\underline{\underline{n}}_\mu = (n_\mu + 1)$ in the case of birth (algorithm 1)

[76] $\underline{\underline{n}}_\mu = (n_\mu - 1)$ in the case of death (algorithm 2)

[77] $\underline{\underline{n}}_\mu = n_\mu$, there is no model transition and the state of the chain is updated using algorithm 3.

[78] Step 2. Propose a transition to the model $GEV_{\underline{\underline{n}}\mu,\underline{\underline{n}}\delta}$ with

[79] $\underline{\underline{n}}_\delta = (n_\delta + 1)$ in the case of birth (algorithm 1′)

[80] $\underline{\underline{n}}_\delta = (n_\delta - 1)$ in the case of death (algorithm 2′)

[81] $\underline{\underline{n}}_\delta = n_\delta$, there is no model transition and the state of the chain is updated using algorithm 3′.

[82] Step 3. Set $n_\mu = \underline{\underline{n}}_\mu$, $n_\delta = \underline{\underline{n}}_\delta$ and return to step 1.

[83] All unchanged dimension parameters are updated using their instrumental distributions. For the shape parameter κ, corresponding transitions are the same as described by *El Adlouni et al.* [2007].

5. Illustrative Examples

[84] To illustrate the proposed approach, two examples are presented here. The first example deals with simulated data. In the second example we study the effect of the Southern Oscillation Index (SOI) on the annual maximum precipitation at the Tehachapi station in California. For both illustrative examples, n_μ(max) = 3 and n_α(max) = 2.

5.1. Simulated Data

[85] A sample of size $n = 50$ is generated from the model $X_t \sim GEV_{2,1}(\mu_t = 100 + 5Y_t, \alpha_t = \exp(0.5), \kappa = -0.1)$, where Y_t is a normal distributed variable with mean 3 and variance 1. Note that hydrometeorological extremes suggest positive skewness, which can be represented by distributions with a right heavy tail. The shape parameter considered for this simulated data corresponds to the coefficient of skewness Cs = 1.86. Figure 3 presents the generated data sample (Figure 3a) and the scatterplot of X_t as a function of the covariate Y_t (Figure 3b). As mentioned in section 2, when the covariate is taken to be different from time, the time series may in fact be stationary.

[86] The sampler was implemented as described in section 4 for the generated data series X_t and the covariate Y_t. The parameters and quantiles are estimated through their empirical posterior distributions. The best model is selected by the proportion of iterations that the MCMC chain remains in that model. For all compared models, the starting values for the BDMCMC are the maximum likelihood estimators of the parameters and the proposal density is Gaussian centered at the current states and has a relatively

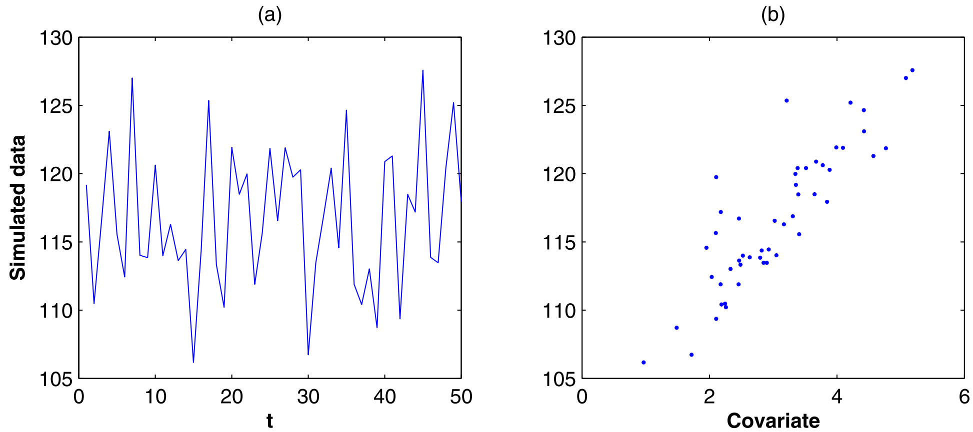

Figure 3. Data series simulated from the $GEV_{2,1}$ model and the normal distribution as covariate.

small variance. The standard deviation of the proposal density is fixed to $10^{-3}|\theta_0|$, where θ_0 is the starting value of any given parameter θ in the BDMCMC algorithm.

[87] A graphical approach is adopted to deal with the convergence of BDMCMC: We run the algorithms for an overly long duration (much longer than is probably needed), and examine the convergence graphs. Figure 4 illustrates the BDMCMC algorithm results with a length of N = 15,000. Figure 4a shows that the parameter n_μ is almost equal to 2 and the parameter n_α transits between 1 and 2 (Figure 4b). The percentage of iterations that the chain remains in each model are 6%, 5%, 51%, 29%, 3%, and 5% for $GEV_{1,1}$, $GEV_{1,2}$, $GEV_{2,1}$, $GEV_{2,2}$, $GEV_{3,1}$, and $GEV_{3,2}$, respectively. Thus the best model corresponds to the $GEV_{2,1}$ model as shown in Figure 4c. The selected model median and the 95% credibility interval are presented in Figure 4d.

[88] Results show that for the selected model the Markov chains, corresponding to the $GEV_{2,1}$ parameters, converge after a small number of iterations (less than 1000). A reasonable burn-in of N_0 = 1000 seems sufficient to guarantee convergence. This is a standard graphical tool to assess the convergence of the Markov chains to their stationary distributions. More discussions on the convergence behavior and the mixing property of the sampler are presented in section 5.2.

[89] The proposal (P2) gives the same results and leads to the same model selection.

5.2. Annual Maximum Precipitation Data

[90] The proposed approach based on the BDMCMC algorithm is used to model the annual maximum precipitation (X (mm)) observed at the Tehachapi station in California (station 048826; latitude, 35.13; longitude, −118.45; period, 1952–2000; sample size n = 49). As this station is located in southern California, its precipitations are strongly affected by the Southern Oscillation Index (SOI). Precip-

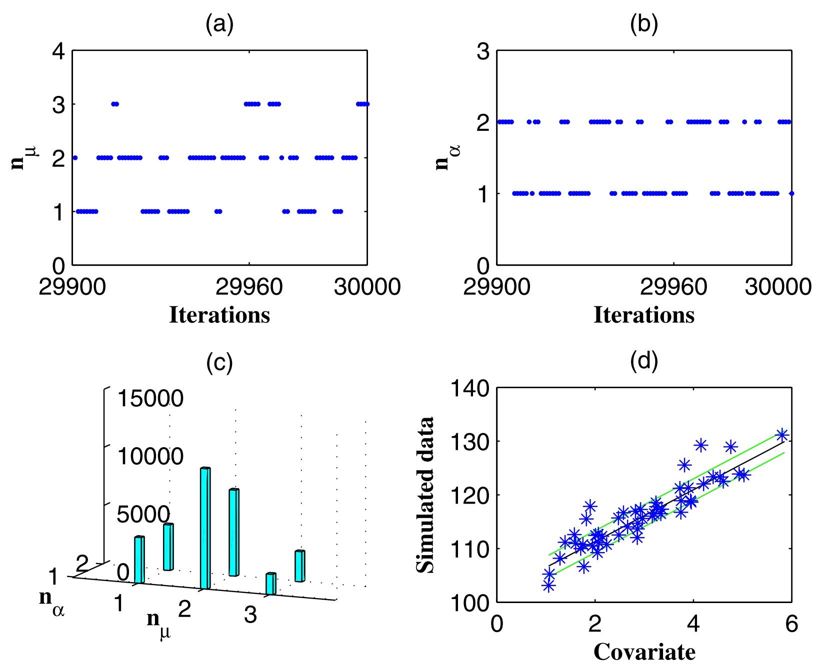

Figure 4. Reversible jump Markov chain Monte Carlo (RJMCMC) algorithm results for the simulated data: (a) MCMC iterations for n_μ, (b) MCMC iterations for n_α, (c) proportions of model selection, and (d) estimated median given by the selected model.

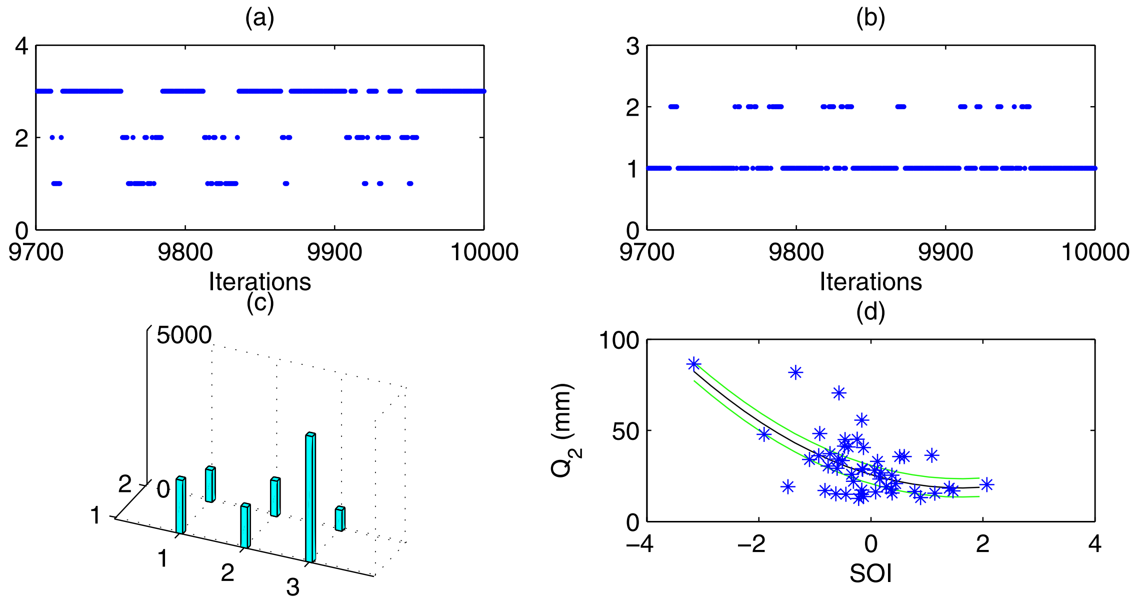

Figure 5. RJMCMC results for the precipitation of the Tehachapi station: (a) MCMC iterations for n_μ, (b) MCMC iterations for n_α, (c) proportions of model selection, and (d) estimated median given by the selected model.

itations at the Tehachapi station were studied by *El Adlouni and Ouarda* [2008], and the deviance information criterion (DIC) was used for model selection. The same problem is considered here in order to illustrate the flexibility of the proposed methodology for parameter estimation and model selection.

[91] The DIC is a simple method to compare the validity of a model M_1 against another model M_0. When M_0 is a special case of M_1 ($M_0 \subset M_1$), the DIC statistic is given by $D = 2\{l_n^*(M_1) - l_n^*(M_0)\}$ [*Coles*, 2001], where $l_n^*(M)$ is the maximized log likelihood function of model M. Large values of D indicate that model M_1 is more adequate and explains more of the data variation than model M_0. The DIC shows that the difference between the $GEV_{1,1}$ and the $GEV_{2,1}$ models is not significant, because $D = 0.95$ is smaller than the 0.95 quantile of the χ_1^2 distribution ($\Pr(\chi_1^2 \leq 0.95) = 0.6703$). However, the DIC shows that the $GEV_{3,1}$ is more adequate than the $GEV_{2,1}$ model. For a more general comparison using DIC, all models should be fitted and compared. This approach is not recommended for extreme model selection [*Coles and Pericchi*, 2003].

[92] The proposed approach, based on the BDMCMC algorithm, allows us to check the adequacy of a more general set of models. Results are presented in Figure 5 and show that the $GEV_{3,1}$ is the most adequate model to represent the dependence between the precipitation at the Tehachapi station and the SOI. The total length of the generated Markov chains for this example is $N = 10{,}000$.

[93] Figures 5a and 5b show that the BDMCMC algorithm is well mixing within the models and that the

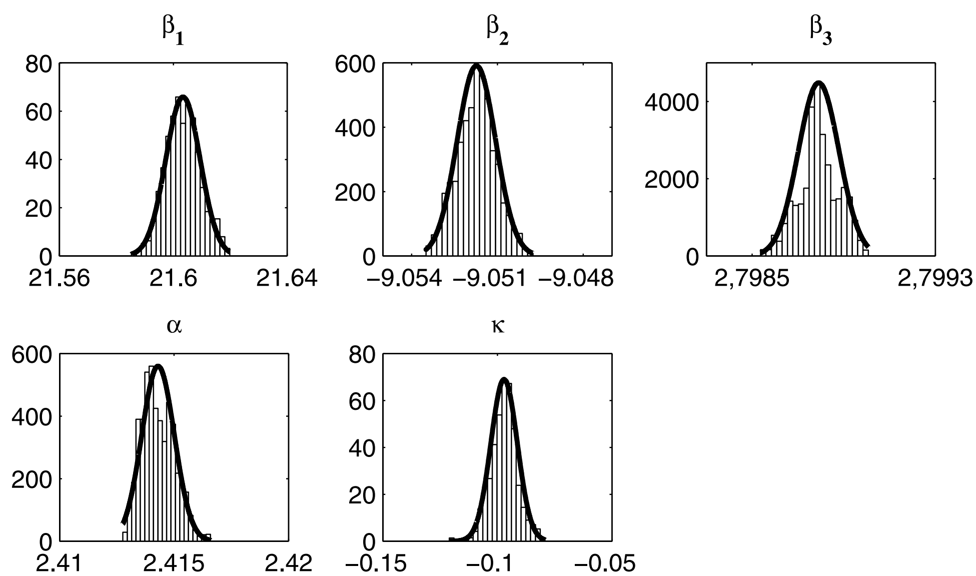

Figure 6. Empirical posterior distribution of the parameter of the $GEV_{3,1}$ model for the Tehachapi example.

Table 1. Estimation of the Annual Maximum Precipitation Median Using the Selected Model $GEV_{3,1}$ and the Classical Model $GEV_{1,1}$ Conditional to Special Values of the Covariate[a]

	SOI = −3.16	SOI = −0.15	SOI = 2.07
$GEV_{1,1}$	27.12	27.12	27.12
$GEV_{3,1}$	87.77	25.72	21.96

[a]Values are in millimeters.

proportion of iterations that the MCMC chain remains in the $GEV_{3,1}$ model is the highest and corresponds to 49% of the length of the generated Markov chains (Figure 5c). Figure 5d presents the estimated annual maximum precipitation medians, conditional to SOI values, with their 95% credibility intervals. Note that the median is simply chosen to illustrate the adequacy of the selected model. On the other hand, all extreme quantiles can be estimated with their credibility intervals.

[94] For convergence assessment, we first examine the mixing over n_μ and n_α, and then the mixing over the GEV with covariate model parameters. To allow the BDMCMC iterations to explore spaces, the starting points should be selected carefully. In a preliminary analysis, results corresponding to starting points $n_\mu = 3$ and $n_\alpha = 1$ show that the sampler is not mixing. The Markov chains remain in the same space given that the likelihood of the $GEV_{3,1}$ model is higher than all competitive models. By choosing, as starting point, the fixed parameter model ($n_\mu = 1$ and $n_\alpha = 1$) the sampler moves quickly from the low likelihood of the starting model to higher likelihood (Figures 5a and 5b). For the rest of the parameters, conditionally on the selected model, the convergence assessment is carried out by graphical tools by observing the variations on the empirical mean and variance. Rigorous diagnosis of MCMC convergence is elusive. Some methods to assess the convergence of MCMC methods, such as Raftery and Lewis and subsampling methods, make it possible to determine the length of the chain and the burn-in time [*Cowles and Carlin*, 1996; *El Adlouni et al.*, 2006].

[95] A burn-in time for this case study is $N_0 = 1000$, and the last 3900 iterations are used to estimate the empirical posterior distribution. Figure 6 shows the histograms of these distributions for the $GEV_{3,1}$ model parameters.

[96] Table 1 presents the annual maximum precipitation median estimated by the selected model $GEV_{3,1}$ and the classical model $GEV_{1,1}$ conditional on the minimum (−3.16), the mean (−0.15), and the maximum (2.07) observed values of the covariate SOI. Results show that the difference between the $GEV_{3,1}$ and the $GEV_{1,1}$ models is larger for negative values of SOI, which correspond to high observed precipitation values. Indeed, the median estimated by the $GEV_{3,1}$ ($Q_{50\%}(GEV_{3,1}) = 87.77$) can be 3 times larger than that estimated by the classic model ($Q_{50\%}(GEV_{1,1}) = 27.12$) for SOI = −3.16. Thus the use of the simplified model could lead to a significant underestimation of the quantiles for some cases of observed SOI values. These results meet those obtained by *El Adlouni and Ouarda* [2008].

6. Conclusions

[97] We have developed a Bayesian approach for both parameter estimation and model selection for the GEV models with covariates based on the birth-death MCMC algorithm. The BDMCMC algorithm was developed to extend the applicability of the MCMC method to problems in which the dimension of the parameter space can change between iterations. The parameter space for such a Markov chain includes the parameters along with an indicator for the current GEV model with covariates.

[98] The proposed procedure can be applied to more general models where parameters are expressed as nonlinear functions of the covariates. Such an approach requires a much longer computation time than the linear case. Also, to obtain summary information on the covariate parameters, large storage space is needed depending on the complexity of the nonlinear functions. However, computational speed is becoming a less determinant factor due to increasing computer power, and the proposed approach can hence be generalized.

[99] As illustrated by simulation results and the applied case study, the Bayesian GEV model with covariates is a valuable tool for parameter varying data analysis of extremes. It has been shown in the present paper that the proposed BDMCMC approach can reliably identify the best models and has performed well with simulated and observed data sets.

[100] **Acknowledgments.** The financial support provided by the National Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chair Program (CRC) is acknowledged. The authors wish also to thank associate editor Philippe Naveau and the three anonymous reviewers whose comments helped considerably improve the quality of the paper.

References

Aitkin, M., R. J. Boys, and T. J. Chadwick (2005), Bayesian point null hypothesis testing via the posterior likelihood ratio, *Stat. Comput.*, *15*(3), 217–230, doi:10.1007/s11222-005-1310-0.

Bayarri, M. J., and J. Berger (2000), P-values for composite null models (with discussion), *J. Am. Stat. Assoc.*, *95*, 1127–1142, doi:10.2307/2669749.

Celeux, G., F. Forbes, C. Robert, and D. M. Titterington (2006), Deviance information criteria for missing data models, *Bayesian Anal.*, *1*(4), 651–674.

Clarke, R. T. (2002), Estimating trends in data from the Weibull and a generalized extreme value distribution, *Water Resour. Res.*, *38*(6), 1089, doi:10.1029/2001WR000575.

Coles, S. G. (2001), *An Introduction to Statistical Modeling of Extreme Values*, 208 pp., Springer, New York.

Coles, S. G., and M. Dixon (1999), Likelihood-based inference for extreme value models, *Extremes*, *2*(1), 5–23, doi:10.1023/A:1009905222644.

Coles, S. G., and L. Pericchi (2003), Anticipating catastrophes through extreme value modeling, *J. R. Stat. Soc., Ser. C*, *52*, 405–416.

Coles, S. G., and J. A. Tawn (1996), Modelling extremes of the rainfall process, *J. R. Stat. Soc., Ser. B*, *58*, 329–347.

Cowles, M. K., and B. P. Carlin (1996), Markov chain Monte Carlo convergence diagnostics: A comparative review, *J. Am. Stat. Assoc.*, *91*, 883–904, doi:10.2307/2291683.

Dellaportas, P., and I. Papageorgiou (2006), Multivariate mixtures of normals with unknown number of components, *Stat. Comput.*, *16*, 57–68, doi:10.1007/s11222-006-5338-6.

El Adlouni, S., and T. B. M. J. Ouarda (2008), Comparaison des méthodes d'estimation des paramètres du modèle GEV non-stationnaire, *Rev. Sci. Eau*, *21*(1), 35–50.

El Adlouni, S., A. C. Favre, and B. Bobée (2006), Comparison of methodologies to assess the convergence of Markov chain Monte Carlo methods, *Comput. Stat. Data Anal.*, *50*(10), 2685–2701, doi:10.1016/j.csda.2005.04.018.

El Adlouni, S., T. B. M. J. Ouarda, X. Zhang, R. Roy, and B. Bobée (2007), Generalized maximum likelihood estimators of the non-stationary GEV model parameters, *Water Resour. Res.*, *43*, W03410, doi:10.1029/2005WR004545.

George, E. I., and R. E. McCulloch (1993), Variable selection via Gibbs sampling, *J. Am. Stat. Assoc.*, *88*, 881–889, doi:10.2307/2290777.

Gilks, W., S. Richardson, and D. Spiegelhalter (1996), *Markov Chain Monte Carlo Methods in Practice*, CRC Press, Boca Raton, Fla.

Green, P. J. (1995), Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, *82*(4), 711–732, doi:10.1093/biomet/82.4.711.

Green, P. J. (2003), Trans-dimensional Markov chain Monte Carlo, in *Highly Structured Stochastic Systems*, edited by P. J. Green, N. L. Hjort, and S. Richardson, chap. 6, p. 179, Oxford Univ. Press, Oxford, U. K.

Hastings, W. (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, *57*, 97–109, doi:10.1093/biomet/57.1.97.

Intergovernmental Panel on Climate Change (IPCC) (2007), *Climate Change: Impacts, Adaptation and Vulnerability—Working Group II Contribution to the Fourth Assessment Report of the IPCC Intergovernmental Panel on Climate Change*, 976 pp., Cambridge Univ. Press, Cambridge, UK.

Jenkinson, A. F. (1955), The frequency distribution of the annual maximum (or minimum) of meteorological elements, *Q. J. R. Meteorol. Soc.*, *81*, 158–171, doi:10.1002/qj.49708134804.

Kharin, V. V., and F. W. Zwiers (2005), Estimating extremes in transient climate change simulations, *J. Clim.*, *18*, 1156–1173.

Leadbetter, M. R., G. Lindgren, and H. Rootzen (1983), *Extremes and Related Properties of Random Sequences and Processes*, 336 pp., Springer, New York.

Martins, E. S., and J. R. Stedinger (2000), Generalized maximum likelihood generalized extreme value quantile estimators for hydrologic data, *Water Resour. Res.*, *36*(3), 737–744, doi:10.1029/1999WR900330.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953), Equations of state calculations by fast computing machines, *J. Chem. Phys.*, *21*, 1087–1092, doi:10.1063/1.1699114.

O'Hagan, A. (2006), *Uncertain Judgements: Eliciting Expert's Probabilities*, 338 pp., John Wiley, Hoboken, N. J.

Park, J. (2005), A simulation-based hyperparameter selection for quantile estimation of the generalized extreme value distribution, *Math. Comput. Simul.*, *70*, 227–234, doi:10.1016/j.matcom.2005.09.003.

Pérez, J. M., and J. O. Berger (2002), Expected posterior prior distributions for model selection, *Biometrika*, *89*(3), 491–511, doi:10.1093/biomet/89.3.491.

Reis, D. S., and J. R. Stedinger (2005), Bayesian MCMC flood frequency analysis with historical information, *J. Hydrol.*, *313*(1–2), 97–116, doi:10.1016/j.jhydrol.2005.02.028.

Ribatet, M., E. Sauquet, J. M. Grésillon, and T. B. M. J. Ouarda (2007), Usefulness of the reversible jump Markov chain Monte Carlo model in regional flood frequency Analysis, *Water Resour. Res.*, *43*, W08403, doi:10.1029/2006WR005525.

Richardson, S., and P. J. Green (1997), On Bayesian analysis of mixture models with an unknown number of components (with discussion), *J. R. Stat. Soc., Ser. B*, *59*, 731–792, doi:10.1111/1467-9868.00095.

Robert, C. P., and G. Casella (2004), *Monte Carlo Statistical Methods*, 2nd ed., Springer, New York.

Robert, C. P., T. Rydén, and D. M. Titterington (2000), Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method, *J. R. Stat. Soc., Ser. B*, *62*, 57–76, doi:10.1111/1467-9868.00219.

Stefanakos, C. N., and G. A. Athanassoulis (2006), Extreme value predictions based on nonstationary time series of wave data, *Environmetrics*, *17*, 25–46, doi:10.1002/env.742.

Stephens, M. (2000), Bayesian analysis of mixture models with an unknown number of components: An alternative to reversible jump methods, *Ann. Stat.*, *28*, 40–74, doi:10.1214/aos/1016120364.

Stephenson, A., and J. Tawn (2004), Bayesian inference for extremes: Accounting for the three extremal types, *Extremes*, *7*(4), 291–307, doi:10.1007/s10687-004-3479-6.

Zhang, J. (2007), Likelihood moment estimation for the generalized Pareto distribution, *Aust. N. Z. J. Stat.*, *49*, 69–77.

Zhang, X., F. W. Zweirs, and G. Li (2004), Monte Carlo experiments on the detection of trends in extreme values, *J. Clim.*, *17*, 1945–1952, doi:10.1175/1520-0442(2004)017<1945:MCEOTD>2.0.CO;2.

S. El Adlouni, Institut National de Statistique et d'Économie Appliquée (INSEA), Avenue allal El Fassi, Madinat AL Irfane, Rabat, 10100, Morocco. (eladlouni@insea.ac.ma)

T. B. M. J. Ouarda, Hydro-Quebec/NSERC Chair in Statistical Hydrology, Canada Research Chair on the Estimation of Hydrometeorological Variables, University of Quebec, INRS-ETE, 490, de la Couronne, ON G1K 9A9, Canada.